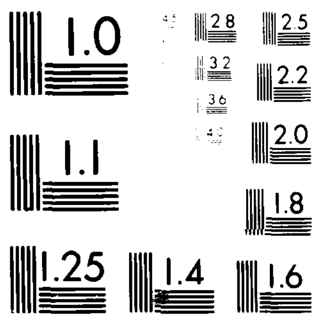
AD-A098 350 NORTH CAROLINA UNIV AT CHAPEL HILL F/G 12/1
FAULTY INSPECTION DISTRIBUTIONS -- SOME GENERALIZATIONS.(U)
APR 81 N L JOHNSON, S KOTZ N00014-81-L-0301
UNCLASSIFIED MIMEOGRAPH SER-1335 NL

1 of 1

END
DATE FILMED
5-81
DTIC



MICROCOPY RESOLUTION TEST CHART

FAULTY INSPECTION DISTRIBUTIONS -- SOME GENERALIZATIONS

by

Norman L. Johnson
University of North Carolina at Chapel Hill

and

Samuel Kotz
University of Maryland, College Park

Abstract

In a previous study

In Johnson, Kotz and Sorkin (1980), the authors derived the distribution of the number of items *observed* to be defective in samples from a finite population, when detection is *erroneous* with a nonzero probability. The present report extends the previous

We extend here the above results by taking into account incorrect identification of nondefectives as well as defectives. Corresponding waiting time distributions are also derived. Furthermore, the case of a stratified finite population corresponding, for example, to defective features of differing severity is considered. Numerical values illustrating the dependence of the corresponding probabilities on the two "misidentification" parameters are presented.

Key Words and Phrases: binomial distribution; compound distributions; faulty identification; hypergeometric distribution; sampling inspection; waiting time; incomplete identification.

1. Introduction

Johnson *et al.* (1980) have discussed some problems arising when attributes inspection is "less than perfect". They considered, in effect, sampling without replacement from a lot of size N containing X defective (or nonconforming) items, when inspection detects such items with probability p. Here this is extended (i) by allowing for a probability, p', of erroneously deciding that an item is defective when really it is not and (ii) by stratifying the population so that inspection error probabilities vary from stratum to stratum.

2. Two Kinds of Inspection Error

The number of defective items, Y, in a random sample (without replacement) of size n has a hypergeometric distribution with parameters n,X,N. Conditionally on Y, the number of items *correctly* called "defective" is distributed as binomial (Y,p) and the number *incorrectly* called "defective" is distributed as binomial (n-Y, p'). Thus, the overall distribution of the total number of items called "defective", Z say, is

$$\mathrm{Bin}(Y,p) + \mathrm{Bin}(n-Y,\ p') \underset{Y}{\wedge} \mathrm{Hypg}(n,X,N)$$

(the two binomial variables being mutually independent), where $\wedge$ denotes the compounding operator (Johnson and Kotz (1969, p. 184)).

Conditional on Y, the r^{th} factorial moment of Z is

$$\mu_{(r)}(Z|Y) = E[Z^{(r)}|Y] = \sum_{j=0}^{r} \binom{r}{j} Y^{(j)} p^{j} (n-Y)^{(r-j)} p'^{\,r-j} .$$

The unconditional r^{th} factorial moment of Z is

$$\mu_{(r)}(Z) = \sum_{j=0}^{r} \binom{r}{j} p^j p'^{\,r-j}\, E[Y^{(j)}(n-Y)^{(r-j)}]$$

$$= \frac{n^{(r)}}{N^{(r)}} \sum_{j=0}^{r} \binom{r}{j} p^j p'^{\,r-j} X^{(j)} (N-X)^{(r-j)} \ . \tag{1}$$

In particular,

$$E[Z] = \{Xp + (N-X)p'\}n/N = n\bar{p} \tag{2.1}$$

$$\mathrm{Var}(Z) = n\bar{p}(1-\bar{p}) - \frac{n(n-1)}{(N-1)} \frac{X}{N} \left(1 - \frac{X}{N}\right)(p-p')^2 \ , \tag{2.2}$$

where $\bar{p} = N^{-1}\{Xp + (N-X)p'\}$. The conditional probability mass function (pmf) of Z given Y is

$$\Pr[Z = z|Y] = \sum_{j=0}^{z} \binom{Y}{j} p^j (1-p)^{Y-j} \binom{n-Y}{z-j} p'^{\,z-j} (1-p')^{n-Y-z+j} \tag{3}$$
$$(z = 0,1,\ldots,Y) \ ,$$

where $\binom{a}{b} = 0$ if $a < b$.

Hence the unconditional pmf of Z is

$$\Pr[Z = z] = \binom{N}{n}^{-1} \sum_{y} \binom{X}{y}\binom{N-X}{n-y} \sum_{j=0}^{z} \binom{y}{j}\binom{n-y}{z-j} p^j (1-p)^{y-j} p'^{\,z-j} (1-p')^{n-y-z+j} \ , \tag{4}$$

when the first $\sum$ is over $\max(0, n-N+X) \le y \le \min(n,X)$.

Numerical values can be obtained expeditiously if adequate tables of hypergeometric probabilities

$$h(y;\, n,X,N) = \binom{N}{n}^{-1} \binom{X}{y}\binom{N-X}{n-y}$$

and binomial probabilities

$$b(y;\, n,p) = \binom{n}{y} p^y (1-p)^{n-y}$$

are available from the formula

$$\Pr[Z = z] = \sum_y h(y;\, n,X,N) \sum_{j=0}^{z} b(j;\, y,p)\, b(z-j;\, n-y,\, p') \ . \qquad (4)'$$

Table 1 gives some examples of distributions, each with sample size n = 10. It is limited by space considerations, but fuller tables have been calculated (for other sample sizes as well as other values of X and N).

p = 0.75

N = 100; X = 5

z\p'	0	0.025	0.05	0.075	0.1
0	.6731	.5244	.4060	.3122	.2384
1	.2818	.3575	.3891	.3901	.3711
2	.0422	.1010	.1608	.2139	.2557
3	.0028	.0156	.0379	.0679	.1028
4	.0001	.0015	.0056	.0138	.0267
5	-	.0001	.0006	.0019	.0047
6	-	-	-	.0002	.0006
7	-	-	-	-	-
8	-	-	-	-	-
9	-	-	-	-	-
10	-	-	-	-	-

N = 200; X = 10

z\p'	0	0.025	0.05	0.075	0.1
0	.6778	.5280	.4087	.3142	.2399
1	.2736	.3520	.3855	.3879	.3698
2	.0446	.1015	.1603	.2129	.2545
3	.0038	.0167	.0388	.0684	.1030
4	.0002	.0017	.0060	.0143	.0272
5	-	.0001	.0006	.0020	.0049
6	-	-	-	.0002	.0006
7	-	-	-	-	.0001
8	-	-	-	-	-
9	-	-	-	-	-
10	-	-	-	-	-

N = 100; X = 10

z\p'	0	0.025	0.05	0.075	0.1
0	.4459	.3488	.2711	.2094	.1606
1	.3892	.3985	.3867	.3611	.3274
2	.1369	.1920	.2378	.2720	.2940
3	.0252	.0513	.0831	.1180	.1532
4	.0027	.0084	.0183	.0327	.0513
5	.0002	.0009	.0027	.0060	.0116
6	-	.0001	.0003	.0008	.0018
7	-	-	-	.0001	.0002
8	-	-	-	-	-
9	-	-	-	-	-
10	-	-	-	-	-

N = 200; X = 20

z\p'	0	0.025	0.05	0.075	0.1
0	.4524	.3537	.2749	.2122	.1627
1	.3803	.3928	.3831	.3591	.3265
2	.1363	.1902	.2354	.2695	.2917
3	.0274	.0529	.0840	.1182	.1529
4	.0034	.0093	.0193	.0336	.0521
5	.0003	.0011	.0030	.0064	.0120
6	-	.0001	.0003	.0008	.0019
7	-	-	-	.0001	.0002
8	-	-	-	-	-
9	-	-	-	-	-
10	-	-	-	-	-

N = 100; X = 20

z\p'	0	0.025	0.05	0.075	0.1
0	.1856	.1465	.1149	.0896	.0694
1	.3540	.3209	.2860	.2510	.2173
2	.2870	.3028	.3093	.3077	.2992
3	.1299	.1619	.1915	.2174	.2386
4	.0362	.0542	.0751	.0980	.1221
5	.0065	.0119	.0195	.0295	.0419
6	.0008	.0017	.0034	.0060	.0098
7	.0001	.0002	.0004	.0008	.0015
8	-	-	-	.0001	.0002
9	-	-	-	-	-
10	-	-	-	-	-

N = 200; X = 40

z\p'	0	0.025	0.05	0.075	0.1
0	.1913	.1509	.1183	.0922	.0714
1	.3507	.3193	.2856	.2514	.2181
2	.2813	.2977	.3051	.3043	.2966
3	.1299	.1609	.1899	.2154	.2364
4	.0382	.0559	.0762	.0987	.1223
5	.0075	.0130	.0206	.0306	.0429
6	.0010	.0021	.0038	.0065	.0103
7	.0001	.0002	.0005	.0009	.0017
8	-	-	-	.0001	.0002
9	-	-	-	-	-
10	-	-	-	-	-

p = 0.8

N = 100; X = 5

z\p'	0	0.025	0.05	0.075	0.1
0	.6545	.5096	.3942	.3029	.2311
1	.2946	.3642	.3915	.3896	.3687
2	.0474	.1073	.1670	.2192	.2597
3	.0034	.0172	.0405	.0712	.1065
4	.0001	.0017	.0062	.0148	.0282
5	-	.0001	.0006	.0021	.0050
6	-	-	-	.0002	.0006
7	-	-	-	-	.0001
8	-	-	-	-	-
9	-	-	-	-	-
10	-	-	-	-	-

N = 200; X = 10

z\p'	0	0.025	0.05	0.075	0.1
0	.6598	.5136	.3973	.3052	.2328
1	.2855	.3581	.3875	.3871	.3673
2	.0499	.1077	.1664	.2180	.2584
3	.0046	.0184	.0414	.0718	.1068
4	.0002	.0020	.0066	.0153	.0287
5	-	.0001	.0007	.0022	.0053
6	-	-	.0001	.0002	.0007
7	-	-	-	-	.0001
8	-	-	-	-	-
9	-	-	-	-	-
10	-	-	-	-	-

N = 100; X = 10

z\p'	0	0.025	0.05	0.075	0.1
0	.4206	.3285	.2549	.1966	.1505
1	.3960	.3991	.3830	.3547	.3196
2	.1500	.2038	.2473	.2788	.2980
3	.0297	.0575	.0905	.1258	.1609
4	.0034	.0099	.0208	.0362	.0557
5	.0002	.0011	.0031	.0069	.0130
6	-	.0001	.0003	.0009	.0020
7	-	-	-	.0001	.0002
8	-	-	-	-	-
9	-	-	-	-	-
10	-	-	-	-	-

N = 200; X = 20

z\p'	0	0.025	0.05	0.075	0.1
0	.4276	.3339	.2591	.1997	.1529
1	.3866	.3930	.3793	.3527	.3186
2	.1490	.2014	.2444	.2759	.2954
3	.0321	.0592	.0913	.1260	.1605
4	.0043	.0110	.0219	.0372	.0566
5	.0004	.0014	.0035	.0074	.0135
6	-	.0001	.0004	.0010	.0022
7	-	-	-	.0001	.0002
8	-	-	-	-	-
9	-	-	-	-	-
10	-	-	-	-	-

N = 100; X = 20

z\p'	0	0.025	0.05	0.075	0.1
0	.1632	.1284	.1004	.0781	.0603
1	.3385	.3038	.2683	.2337	.2009
2	.2980	.3089	.3109	.3054	.2937
3	.1463	.1775	.2056	.2295	.2482
4	.0442	.0638	.0860	.1098	.1343
5	.0086	.0150	.0237	.0349	.0486
6	.0011	.0023	.0044	.0075	.0119
7	.0001	.0002	.0005	.0011	.0020
8	-	-	-	.0001	.0002
9	-	-	-	-	-
10	-	-	-	-	-

N = 200; X = 40

z\p'	0	0.025	0.05	0.075	0.1
0	.1691	.1330	.1039	.0807	.0623
1	.3359	.3028	.2684	.2344	.2020
2	.2916	.3033	.3063	.3018	.2910
3	.1457	.1760	.2035	.2269	.2455
4	.0464	.0655	.0871	.1103	.1343
5	.0098	.0163	.0251	.0362	.0498
6	.0014	.0028	.0049	.0082	.0127
7	.0001	.0003	.0007	.0012	.0022
8	-	-	.0001	.0001	.0002
9	-	-	-	-	-
10	-	-	-	-	-

p = 0.85

z\φ'	N = 100; X = 5					N = 200; X = 10				
	0	0.025	0.05	0.075	0.1	0	0.025	0.05	0.075	0.1
0	.6363	.4950	.3827	.2938	.2240	.6421	.4995	.3861	.2964	.2259
1	.3068	.3705	.3936	.3889	.3661	.2968	.3637	.3892	.3862	.3646
2	.0528	.1136	.1732	.2244	.2637	.0553	.1140	.1724	.2230	.2621
3	.0040	.0189	.0431	.0746	.1103	.0054	.0203	.0442	.0753	.1106
4	.0001	.0019	.0067	.0158	.0298	.0003	.0023	.0073	.0164	.0303
5	-	.0001	.0007	.0023	.0054	-	.0002	.0008	.0024	.0057
6	-	-	-	.0002	.0007	-	-	.0001	.0002	.0007
7	-	-	-	-	.0001	-	-	-	-	.0001
8	-	-	-	-	-	-	-	-	-	-
9	-	-	-	-	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-	-	-
	N = 100; X = 10					N = 200; X = 20				
0	.3964	.3091	.2396	.1844	.1410	.4040	.3150	.2441	.1879	.1436
1	.4012	.3985	.3785	.3479	.3115	.3914	.3922	.3747	.3459	.3106
2	.1632	.2153	.2563	.2851	.3015	.1615	.2123	.2529	.2817	.2986
3	.0347	.0641	.0980	.1337	.1686	.0373	.0658	.0988	.1338	.1680
4	.0042	.0116	.0235	.0398	.0603	.0053	.0129	.0248	.0410	.0613
5	.0003	.0013	.0037	.0079	.0145	.0005	.0017	.0042	.0085	.0151
6	-	.0001	.0004	.0010	.0024	-	.0001	.0005	.0012	.0026
7	-	-	-	.0001	.0003	-	-	-	.0001	.0003
8	-	-	-	-	-	-	-	-	-	-
9	-	-	-	-	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-	-	-
	N = 100; X = 20					N = 200; X = 40				
0	.1432	.1123	.0876	.0678	.0522	.1492	.1170	.0912	.0706	.0543
1	.3219	.2861	.2507	.2166	.1849	.3199	.2857	.2512	.2178	.1863
2	.3066	.3127	.3105	.3014	.2868	.2996	.3067	.3056	.2977	.2841
3	.1626	.1927	.2190	.2405	.2565	.1613	.1905	.2162	.2373	.2533
4	.0530	.0741	.0974	.1219	.1466	.0553	.0758	.0984	.1222	.1463
5	.0110	.0186	.0285	.0410	.0559	.0126	.0202	.0301	.0425	.0572
6	.0015	.0031	.0056	.0093	.0144	.0019	.0036	.0063	.0101	.0153
7	.0001	.0003	.0007	.0014	.0025	.0002	.0004	.0009	.0016	.0028
8	-	-	.0001	.0001	.0003	-	-	.0001	.0002	.0003
9	-	-	-	-	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-	-	-

$\underline{p = 0.9}$

	N = 100; X = 5					N = 200; X = 10				
z\p'	0	0.025	0.05	0.075	0.1	0	0.025	0.05	0.075	0.1
0	.6184	.4808	.3714	.2849	.2170	.6249	.4858	.3752	.2878	.2192
1	.3183	.3762	.3953	.3879	.3634	.3074	.3689	.3905	.3849	.3618
2	.0584	.1201	.1793	.2296	.2675	.0610	.1203	.1783	.2279	.2657
3	.0047	.0206	.0458	.0780	.1141	.0063	.0222	.0470	.0787	.1144
4	.0002	.0021	.0073	.0169	.0313	.0004	.0026	.0080	.0176	.0320
5	-	.0001	.0008	.0024	.0058	-	.0002	.0009	.0027	.0061
6	-	-	.0001	.0002	.0007	-	-	.0001	.0003	.0008
7	-	-	-	-	.0001	-	-	-	-	.0001
8	-	-	-	-	-	-	-	-	-	-
9	-	-	-	-	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-	-	-

	N = 100; X = 10					N = 200; X = 20				
0	.3733	.2907	.2250	.1729	.1319	.3815	.2971	.2298	.1766	.1347
1	.4049	.3968	.3734	.3407	.3032	.3947	.3904	.3696	.3287	.3024
2	.1762	.2263	.2648	.2908	.3045	.1738	.2227	.2609	.2870	.3012
3	.0400	.0709	.1057	.1417	.1762	.0428	.0726	.1064	.1415	.1754
4	.0052	.0135	.0263	.0437	.0651	.0065	.0150	.0278	.0450	.0661
5	.0004	.0016	.0043	.0089	.0161	.0006	.0020	.0049	.0096	.0169
6	-	.0001	.0005	.0012	.0027	-	.0002	.0006	.0014	.0030
7	-	-	-	.0001	.0003	-	-	-	.0001	.0003
8	-	-	-	-	-	-	-	-	-	-
9	-	-	-	-	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-	-	-

	N = 100; X = 20					N = 200; X = 40				
0	.1253	.0980	.0761	.0587	.0450	.1314	.1027	.0798	.0615	.0471
1	.3044	.2682	.2331	.1999	.1695	.3031	.2684	.2341	.2015	.1713
2	.3128	.3142	.3081	.2957	.2786	.3052	.3079	.3031	.2920	.2760
3	.1787	.2073	.2314	.2503	.2635	.1766	.2043	.2279	.2465	.2598
4	.0626	.0851	.1093	.1342	.1588	.0649	.0867	.1101	.1343	.1582
5	.0140	.0227	.0339	.0477	.0638	.0159	.0246	.0358	.0493	.0652
6	.0020	.0040	.0070	.0114	.0173	.0026	.0047	.0079	.0124	.0184
7	.0002	.0005	.0010	.0018	.0031	.0003	.0006	.0012	.0021	.0035
8	-	-	.0001	.0002	.0004	-	-	.0001	.0002	.0004
9	-	-	-	-	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-	-	-

p = 0.95

z\p'	N = 100; X = 5					N = 200; X = 10				
	0	0.025	0.05	0.075	0.1	0	0.025	0.05	0.075	0.1
0	.6009	.4668	.3603	.2762	.2102	.6080	.4723	.3645	.2794	.2126
1	.3291	.3816	.3967	.3867	.3606	.3174	.3736	.3916	.3835	.3588
2	.0642	.1266	.1854	.2346	.2712	.0668	.1267	.1841	.2327	.2691
3	.0055	.0225	.0486	.0815	.1180	.0074	.0242	.0500	.0823	.1182
4	.0002	.0024	.0080	.0180	.0330	.0005	.0029	.0087	.0188	.0337
5	-	.0002	.0009	.0026	.0062	-	.0002	.0010	.0029	.0065
6	-	-	.0001	.0003	.0008	-	-	.0001	.0003	.0009
7	-	-	-	-	.0001	-	-	-	-	.0001
8	-	-	-	-	-	-	-	-	-	-
9	-	-	-	-	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-	-	-

z\p'	N = 100; X = 10					N = 200; X = 20				
0	.3514	.2732	.2111	.1619	.1234	.3601	.2800	.2163	.1659	.1264
1	.4071	.3942	.3676	.3331	.2947	.3967	.3876	.3638	.3312	.2941
2	.1890	.2369	.2728	.2959	.3069	.1858	.2327	.2682	.2916	.3033
3	.0457	.0780	.1136	.1497	.1837	.0486	.0797	.1142	.1493	.1826
4	.0063	.0155	.0294	.0478	.0701	.0078	.0173	.0311	.0492	.0711
5	.0005	.0020	.0050	.0101	.0178	.0008	.0025	.0057	.0109	.0187
6	-	.0002	.0006	.0014	.0031	.0001	.0002	.0007	.0017	.0034
7	-	-	-	.0001	.0004	-	-	.0001	.0002	.0004
8	-	-	-	-	-	-	-	-	-	-
9	-	-	-	-	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-	-	-

z\p'	N = 100; X = 20					N = 200; X = 40				
0	.1093	.0852	.0659	.0507	.0387	.1155	.0900	.0696	.0535	.0409
1	.2863	.2502	.2157	.1838	.1547	.2858	.2510	.2174	.1858	.1569
2	.3166	.3137	.3039	.2887	.2694	.3086	.3072	.2989	.2850	.2669
3	.1943	.2210	.2427	.2588	.2692	.1912	.2171	.2384	.2545	.2649
4	.0730	.0967	.1216	.1467	.1710	.0753	.0981	.1221	.1464	.1700
5	.0175	.0274	.0400	.0549	.0722	.0197	.0296	.0420	.0567	.0736
6	.0027	.0051	.0087	.0138	.0205	.0034	.0060	.0098	.0150	.0218
7	.0003	.0006	.0012	.0023	.0039	.0004	.0008	.0015	.0027	.0044
8	-	-	.0001	.0002	.0005	-	.0001	.0002	.0003	.0006
9	-	-	-	-	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-	-	-

Note that for $p' = 0$, the same distribution is obtained if the values of n and X are interchanged.

As N and X are increased proportionately to each other with $X/N = \omega$, say, the other parameters (n,p,p') remaining constant, the distribution of Z tends to a binomial with parameters n, $XN^{-1}p + (1 - XN^{-1})p'$.

Another simple special case is $p = p'$, leading to a binomial distribution with parameters n,p (whatever the values of X and N). However, this is a most unlikely situation -- it would correspond to completely useless inspection, unable to differentiate between satisfactory and nonsatisfactory items.

The distributions shown are quite sensitive to the value of p' (false condemnation) because the ratio X/N is relatively small. When the proportion of nondefectives $(1 - X/N)$ is lower, p' has less effect.

3. Stratified Populations

More generally, we can suppose the lot divided into k strata $\pi_1,\pi_2,\ldots,\pi_k$ of sizes $N_1,N_2,\ldots,N_k$ $(\sum_{j=1}^k N_j = N)$ such that for any chosen individual in π_j, the probability of "detection as defective" (whether this is really so or not) is p_j. The different strata may, for example, correspond to actual defects of differing degrees of visibility. The case considered in Section 2 corresponds to $k = 2$, $p_1 = p$, $p_2 = p'$, $N_1 = X$, $N_2 = N - X$.

The number observed as "defective" in a random sample of size n is then distributed as

$$Z \frown \sum_{j=1}^{k} \text{Bin}(Y_j,p_j) \bigwedge_{Y} \text{Mult Hypg}_k(n; N_1,\ldots,N_k; N) \,. \tag{5}$$

The binomials are mutually independent, conditional on $\underset{\sim}{Y}$; for the multivariate hypergeometric, $\Pr[\underset{\sim}{Y} = \underset{\sim}{y}] = \binom{N}{n}^{-1} \prod_{j=1}^{k} \binom{N_j}{y_j}$ $(\sum_{j=1}^{k} y_j = n)$. Then

$$E[Z^{(r)}|\underset{\sim}{Y}] = \sum_r' \frac{r!}{\prod_{j=1}^{k} r_j!} \prod_{j=1}^{k} (p_j^{r_j} Y_j^{(r_j)}) \ , \tag{6}$$

where $\sum_r'$ denotes summation over nonnegative integers $r_1, \ldots, r_k$ such that $\sum_{j=1}^{k} r_j = r$. Taking expectations with respect to $\underset{\sim}{Y}$,

$$E[Z^{(r)}] = \sum_r' \frac{r!}{\prod_{j=1}^{r} r_j!} \prod_{j=1}^{k} p_j^{r_j} \frac{n^{(r)}}{N^{(r)}} N_j^{(r_j)}$$

$$= \frac{n^{(r)} r!}{N^{(r)}} \sum_r' \prod_{j=1}^{k} \left(\frac{N_j^{(r_j)} p_j^{r_j}}{r_j!} \right) . \tag{7}$$

In particular,

$$E[Z] = \frac{n}{N} \sum_{j=1}^{k} N_j p_j = n\bar{p} \tag{8.1}$$

and

$$E[Z(Z-1)] = \frac{n(n-1)}{N(N-1)} [\sum_{j=1}^{k} N_j(N_j-1)p_j^2 + 2 \sum\sum_{j<j'}^{k} N_j N_{j'} p_j p_{j'}]$$

$$= \frac{n(n-1)}{N(N-1)} [(\sum_{j=1}^{k} N_j p_j)^2 - \sum_{j=1}^{k} N_j p_j^2] \ , \tag{8.2}$$

whence

$$\mathrm{Var}(Z) = n\bar{p}(1-\bar{p}) - \frac{n(n-1)}{N(N-1)} \sum_{j=1}^{k} N_j(p_j-\bar{p})^2 \tag{8.3}$$

where $\bar{p} = N^{-1} \sum_{j=1}^{k} N_j p_j$.

Alternative and instructive formulae for the variance are

$$\mathrm{Var}(Z) = \frac{n(N-n)}{N-1}\, \bar{p}(1-\bar{p}) + \frac{n(n-1)}{N(N-1)}(\bar{p} - N^{-1} \sum_{j=1}^{k} N_j p_j^2) \qquad (8.3)'$$

$$\mathrm{Var}(Z) = \frac{n(N-n)}{N-1}\, \bar{p}(1-\bar{p}) + \frac{n(n-1)}{N^2(N-1)} \sum_{j=1}^{k} N_j p_j (1-p_j)\ . \qquad (8.3)''$$

The first term in (8.3)' and (8.3)" corresponds to the variance of the (actual) number of defectives in a random sample (without replacement) of size n from a population of size N containing $N\bar{p}$ defectives. It follows that the variance of Z is not less than this, while from (8.3) it cannot exceed $n\bar{p}(1-\bar{p})$ -- the value it would have in sampling with replacement from the same population (when the distribution of Z would be binomial with parameters $n,\bar{p}$). This will, of course, also be a good approximation when N is large.

As a limiting case, we might have $N_j = 1$ and $k = N$ -- that is, each item in the lot would have its own probability (p_j) of being declared defective.

Note that this differs from a model in which there is supposed to be a prior distribution of the probability of being declared defective, and the p_j's are regarded as realized values from this distribution. For this latter model, we reach, in effect, a "with replacement" situation, as distinguished from the "without replacement" model we have considered, with the distribution of Z depending on the specific values of the p_j's.

The "with replacement" model can be regarded as a mixture of "without replacement" models -- the latter being conditional on the specific sets of values $p_1, p_2, \ldots, p_N$. The average variance of the number of items declared defective for the "without replacement" will, in general, be smaller than that for the "with replacement" -- because the latter is increased by variation among the p's. Formally,

$$\mathrm{Var}(Z) = E_{\underset{\sim}{p}}[\mathrm{Var}(Z)\underset{\sim}{p})] + \mathrm{Var}_{\underset{\sim}{p}}(E[Z|\underset{\sim}{p}]) \ . \tag{9}$$

4. A Waiting Time Distribution

Suppose now that, under the conditions of Section 3, items are inspected one at a time (without replacement) until a predetermined number a of items have been assessed as "defective". Denoting by M the number of items needed to attain this goal, we have (cf. Section 4 of Johnson *et al.* (1980))

$$\Pr(M>m) = \Pr(Z<a) \qquad m = 1,2,\ldots,N-1 \qquad (a \le N) \ , \tag{10}$$

where Z has the distribution (5) with n replaced by m. Note that in this case $\underline{a}$ can exceed the actual number of defectives in the lot because an item can be assessed as defective even though it is not.

The distribution of M is not proper because there is a positive probability that even when all N items in the lot have been inspected, fewer than $\underline{a}$ items will be declared "defective".

In order to derive the distribution of M, we first consider the possible sets of decisions for the N items. The distribution of the number (D) of those which will be "defective" is that of the sum of independent binomial variables $B_1, B_2, \ldots, B_k$ with parameters

$(N_1,p_1),(N_2,p_2),\ldots,(N_k,p_k)$ respectively. Given D, each of the $\binom{N}{D}$ possible orderings of the D "defective" and (N-D) "not defective" decisions is equally likely, and the number of items up to and including the a^{th} defective (M) has the negative hypergeometric distribution

$$\Pr(M=m) = \binom{N}{D}^{-1} \binom{m-1}{a-1}\binom{N-m}{D-a} \qquad (m = a, a+1,\ldots, N-D+a) \; , \tag{11}$$

provided $D \geq a$.

The conditional expected value of M given D ($\geq a$) is

$$E[M|D] = a(N+1)/(D+1) \; . \tag{12}$$

If we neglect the possibility that D is less than a, the overall expected value of M is approximately

$$\begin{aligned} a(N+1)E[(D+1)^{-1}] &\doteq [\{E[D] + 1\}^{-1} + \{E[D] + 1\}^{-3} \operatorname{Var}(D)](N+1)a \\ &= a(N+1)(\sum_{j=1}^{k} N_j p_j + 1)^{-1} [1 + (\sum_{j=1}^{k} N_j p_j + 1)^{-2} \sum_{j=1}^{k} N_j p_j (1-p_j)] \\ &\doteq a\bar{p}^{-1}[1 + \frac{1}{N}\{1 + \frac{1}{\bar{p}} + \frac{1}{\bar{p}^2} \sum_{j=1}^{k} \omega_j p_j (1-p_j)\}] \; , \end{aligned} \tag{13}$$

where $\omega_j = N_j/N$ (the proportion in the j^{th} stratum).

We evaluate the overall variance of M as

$$\begin{aligned} \operatorname{Var}(M) &= \operatorname{Var}(E[M|D]) + E[\operatorname{Var}(M|D)] \qquad (\text{cf. } (9)) \\ &= a^2(N+1)^2 \operatorname{Var}((D+1)^{-2}) + a(N+1)E\left[\frac{(N-D)(D+1-a)}{(D+1)^2 \, (D+2)}\right] . \end{aligned}$$

After straightforward though quite tedious calculations, we find

$$\mathrm{Var}(M) \doteq \frac{a^2}{N\bar{p}^4} \sum_{j=1}^{k} \omega_j p_j (1-p_j)$$

$$+ \frac{a(1-\bar{p})}{\bar{p}^2} [1 + \frac{1}{N} + \frac{1}{N\bar{p}} \left\{ \frac{3-\bar{p}}{\bar{p}(1-\bar{p})} \sum_{j=1}^{k} \omega_j p_j (1-p_j) - (3+a) \right\}]$$

$$\doteq \frac{a(1-\bar{p})}{\bar{p}^2} [1 + \frac{1}{N} - \frac{3+a}{N\bar{p}} + \frac{3+a-(a+1)\bar{p}}{N\bar{p}^2(1-\bar{p})} \sum_{j=1}^{k} \omega_j p_j (1-p_j)] \ . \qquad (14)$$

As $N \to \infty$, the expected value of M tends to $a\bar{p}^{-1}$ and the variance tends to $a\bar{p}^{-2}(1-\bar{p})$. These are the mean and variance, respectively, of the (negative binomial) waiting time distribution for occurrence of <u>a</u> "successes" in independent trials with probability of success equal to $\bar{p}$ at each trial.

(14) can be written

$$\mathrm{Var}(M) \doteq \frac{a(1-\bar{p})}{\bar{p}^2} \left[1 + \frac{1}{N} - \frac{3+a}{N\bar{p}} \left\{ 1 - \frac{\sum \omega_j p_j (1-p_j)}{\bar{p}(1-\bar{p})} \right\} - \frac{a+1}{N} \cdot \frac{\sum \omega_j p_j (1-p_j)}{\bar{p}(1-\bar{p})} \right] .$$

Since $\sum_{j=1}^{k} \omega_j p_j (1-p_j) = \bar{p} - \sum_{j=1}^{k} \omega_j p_j^2 \le \bar{p} - \bar{p}^2 = \bar{p}(1-\bar{p})$ and $\frac{a+3}{\bar{p}} > a + 1$, it follows that

$$\frac{a(1-\bar{p})}{\bar{p}^2} [1 + \frac{1}{N} - \frac{3+a}{N\bar{p}}] \le \mathrm{Var}(M) \le \frac{a(1-\bar{p})}{\bar{p}^2} (1 + \frac{1}{N}) \ .$$

<u>Acknowledgement</u>

Samuel Kotz's work was supported by the U.S. Office of Naval Research under Contract N00014-81-K-0301.

References

Johnson, N.L. and Kotz, S. (1969). *Distributions in Statistics -- Discrete Distributions*. New York: John Wiley and Sons.

Johnson, N.L., Kotz, S. and Sorkin, H.L. (1980). Faulty inspection distributions. *Comm. Statist. A9*, 917-922.

Sorkin, H.L. (1977). An Empirical Study of Three Confirmation Techniques: Desirability of Expanding the Respondent's Decision Field. Ph.D. Thesis, University of Minnesota.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) UNCLASSIFIED

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-A098 350	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Faulty Inspection Distributions -- Some Generalizations.		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL rept.
		6. PERFORMING ORG. REPORT NUMBER Mimeo Series No. 1335
7. AUTHOR(s) Norman L. Johnson and Samuel Kotz		8. CONTRACT OR GRANT NUMBER(s) ONR Contract N00014-81-K-0301
9. PERFORMING ORGANIZATION NAME AND ADDRESS		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Office of Naval Research Statistics and Probability Program (Code 436) Arlington, VA 22217		12. REPORT DATE April 1981
		13. NUMBER OF PAGES 16
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) MIMEOGRAPH, SER-1335		15. SECURITY CLASS (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release -- Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

binomial distribution; compound distributions; faulty identification; hypergeometric distribution; sampling inspection; waiting time; incomplete identification.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In Johnson, Kotz and Sorkin (1980), the authors derived the distribution of the number of items *observed* to be defective in samples from a finite population, when detection is *erroneous* with a nonzero probability.

We extend here the above results by taking into account incorrect identification of nondefectives as well as defectives. Corresponding waiting time distributions are also derived. Furthermore, the case of a stratified finite population corresponding, for example, to defective

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

357300

20\. features of differing severity is considered. Numerical values illustrating the dependence of the corresponding probabilities on the two "misidentification" parameters are presented.

DATE
ILMED
-8